

<https://helda.helsinki.fi>

Temporal boundary of auditory event formation : An electrophysiological marker

Simon, Júlia

2019-06

Simon , J , Balla , V & Winkler , I 2019 , ' Temporal boundary of auditory event formation : An electrophysiological marker ' , International Journal of Psychophysiology , vol. 140 , pp. 53-61 . <https://doi.org/10.1016/j.ijpsycho.2019.04.006>

<http://hdl.handle.net/10138/325021>

<https://doi.org/10.1016/j.ijpsycho.2019.04.006>

cc_by_nc_nd

acceptedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

1
2
3 Title:
4

5
6 Temporal boundary of auditory event formation: An electrophysiological marker
7
8
9

10 Júlia Simon (corresponding author)
11

12
13 Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural
14 Sciences, Hungarian Academy of Sciences, Budapest, Hungary; Department of
15 Cognitive Science, Faculty of Natural Sciences, Budapest University of Technology
16 and Economics, Budapest, Hungary
17
18

19
20 Contact information: e-mail: simon.julia@ttk.mta.hu, simon.julia89@gmail.com
21

22 Address: 1117 Budapest, Magyar tudósok körútja 2.
23

24
25 Phone number: +36 1 382 6809, +36 30 9237539
26

27 Viktória Balla
28

29
30 University of Szeged, Institute of Psychology, Szeged, Hungary; Cognitive Brain
31 Research Unit, Department of Psychology and Logopedics, Faculty of Medicine,
32 University of Helsinki, Helsinki, Finland
33
34

35 E-mail: ballaviktoria93@gmail.com
36

37 István Winkler
38

39
40 Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural
41 Sciences, Hungarian Academy of Sciences, Budapest, Hungary;
42

43 E-mail: winkler.istvan@ttk.mta.hu
44
45
46
47

48 Declarations of interests: none.
49
50
51
52
53
54
55
56

Abstract

The formation of auditory events requires integration between successive sounds. There is a temporal limit below which a single sound event is perceived while above which a second perceptual event is formed. Behavioral studies applying the Temporal Order Judgment paradigm showed that this boundary is between 20 and 70 ms. Here we provide event-related potential (ERP) evidence from two experiments showing a qualitative change in the processing of tone pairs between 25 and 75 ms within-pair inter-stimulus intervals (ISI). We also show that this temporal boundary can be influenced by the immediate acoustical context, the statistical distribution of the ISIs within the sequence of tone-pairs.

Keywords: temporal order judgment (TOJ), temporal integration, ERP, perceptual event formation

1. Introduction

Temporal changes provide essential information for cognition. They are especially crucial for auditory perception, which substantially relies on sequential decoding of the incoming stimuli in order to identify objects. Temporal processing in the human brain can be studied on different time scales from microsecond resolution allowing sound localization to circadian rhythms of multiple hours influencing sleep-wake cycles (Mauk & Buonomano, 2004). The present study focuses on the temporal resolution of the auditory system determining the limits of perceptual event formation, the formation of information chunks organized in so called perceptual moments (Pöppel, 1997; Stroud, 1956; Kristofferson, 1967; Shallice, 1967), because it directly influences cognition and

motor organization (Pöppel, 1997, 2004; Humes et al., 2013). The length of the perceptual moment is primarily estimated through psychophysical methods whose results may also depend on higher level processes. Therefore, we looked for an electrophysiological marker/index of the lower temporal limit of the formation of a second perceptual moment, a qualitative change in the electrophysiological response as a function of the interval between two sounds (“electrophysiological boundary”) that would be resistant to individual response strategies. Such a marker could be utilized for better understanding the properties of auditory event formation, such as its sensitivity to contextual changes. In two experiments, we explored these issues.

‘*Perceptual moment*’ is a block of information defined by a time interval that has only one time tag within the auditory system and therefore, the temporal succession of its *constituent physical events* cannot be reliably established. The content of a moment, a unitary *perceptual event* can reach consciousness although this is not necessary. Note that a perceptual event is a result of a cognitive analysis while a physical event is defined through theoretical (mathematical) separability of the stimulus (e.g., two tones separated by a silent period). Pöppel (1997) suggested that the length of the perceptual moment can be assessed by temporal order judgments (TOJ). Based mainly on such measurements, the length of a perceptual moment has been estimated to be ca. 20-70 ms (Fostick & Babkoff, 2013; VanRullen & Koch, 2003; Szymaszek et al., 2009; Pöppel, 1971, 1997, 2004; Hirsh & Sherrick, 1961; von Baer, 1864). The accuracy of order judgements decreases with decreasing inter stimulus intervals (ISI) or with stimulus onset asynchrony (SOA). The distinction between ISI and SOA is not practically relevant with non-overlapping stimuli having short durations. However, the loss of order

detection is not equal to the loss of correct cardinality perception (the perception of one vs. two sounds); indeed, synchrony judgments are dissociable from order judgments (Binder, 2015; Love et al., 2013; Van Eijk et al. 2008). The number of sounds can be correctly judged earlier (after 5-10 ms, Litovsky et al., 1999) than their order, therefore it is easier to judge non-synchrony than the correct order. With the goal of finding a physiological correlate of the perceptual moment, the current study measured temporal order judgments.

The grouping of continuous or successive input signals into a unitary perceptual chunk accompanied by the loss of temporal order has been referred to as integration or fusion in the literature. However these concepts have been applied to very different processes. Here we attempt to clarify the concept of the grouping process assumed to underlie perceptual event formation by contrasting it with other types of grouping processes.

In the research of sound perception, the term integration usually refers to the interdependence of consecutive events, while the term ‘temporal integration window’ refers to a 150-200 ms interval within which sound energy is summed and constituent events (if any) are strongly associated (Zwislocki et al., 1960; Pedersen & Salomon, 1977; Saija et al., 2014). Violating one of the constituent events within this interval elicits the Mismatch Negativity (MMN; Näätänen, 1990) event-related brain potential (ERP) response (Yabe et al., 1997; Yabe et al., 1998; Sussman et al., 1999). This interval also matches the temporal extent of the “short auditory store” proposed by Cowan (1984), which can be considered as a process rather than a static store (for compatible concepts, see Massaro, 1975; Näätänen & Winkler, 1999). The temporal integration

window is known to be sensitive to spectral grouping, such that probably multiple temporal windows coexist for the same period, covering widely different spectral ranges of the incoming stimulus. This explains some auditory streaming phenomena (see, e.g., Shinozaki et al., 2003) as well as the lack of masking within continuous speech (noted by Cowan, 1984). In contrast, the processes associated with perceptual moment formation, are less dependent on spectral features (see, e.g., Simon and Winkler, 2018). Further, the minimal interval for perceptual moment formation is much shorter than the temporal integration window (20-70 ms as opposed to 150-250 ms); thus they probably represent two different time-scales within the temporal structure of perception (see, e.g., Pöppel, 1997). The processes forming perceptual moments are often referred to as fusion (a strong merging of events) rather than integration. However, fusion is a concept generally used by the precedence effect phenomena (Litovsky et al., 1999), the collection of processes requiring shorter onset asynchronies than perceptual moment formation. The precedence effect refers to the dominance of the leading sound. If the onset asynchrony is below ~1 ms, a unitary percept is formed from the two sounds and, although auditory features of the event, such as its perceived location are influenced more by the leading sound, the lagging sound also affects the percept (e.g., a left- and a right-side sound together is localized to the center). At onset asynchronies around 5-10 ms, an echo threshold is found with the fused percept splitting into a percept with two components. This threshold is more or less universal among vertebrate species (Dent & Dooling, 2004). However, longer thresholds have also been reported with speech sounds (Haas, 1951). As a result, Cutting (1976) differentiated six types of fusions in dichotic listening to linguistic stimuli.

In the current study, we focus on the temporal boundary at which the unitary perceptual event, characterized by the impression of a coherent double sound (“integrated pair”), splits into two consecutive *perceptual events*, each having ‘time tags’ that allow correct order judgement. Conceptualizing the temporal resolution of perceptual event formation by order discrimination has been also used in clinical applications, as temporal order training was shown lead to performance improvements in patients with temporal lesions (Steinbüchel & Pöppel, 1993) and with developmental language disorders (Fostick et al., 2014; Tallal et al., 1996; Merzenich, 1996).

The formation of perceptual units has been proposed to be implemented in the brain by high-frequency oscillations (Schwender et al., 1994; Pöppel, 1997), such as the gamma oscillation (Galambos et al., 1981) or by the half cycle of the alpha rhythm (Kristofferson, 1967). In contrast to a pure integration process that can be conceptualized as a ‘neural high pass filter’, discretization of the input by neural oscillations implies that perceptual event formation is jointly determined by the temporal separation between consecutive constituent events and the phase of these events with respect to an intrinsic neuronal oscillatory process (VanRullen & Koch, 2003). The notion of discretization by neural oscillations is gaining interest again in recent research (VanRullen & Koch, 2003; Poeppel, 2003; Cecere et al., 2014; Florin et al., 2017). However, the separation of consecutive events does not necessarily require neural oscillations; it might also be the result of difference computation between changing system states (Mauk & Buonomano, 2004). We note that the current study is neutral regarding the proposed mechanisms of event separation.

Few electroencephalographic (EEG) studies employed stimuli compatible with the spatial version of the temporal order judgment (TOJ) task, which requires listeners to judge the order of two sounds differing in their source location. There is also a spectral version of the TOJ task, which requires listeners to judge the order of two tones of different pitch. However, recent results suggest that performance in the spectral TOJ task is based on a holistic processing strategy, rather than pure order judgement (Fink, Churan & Wittmann, 2005; Szymaszek et al., 2009; Fostick & Babkoff, 2013). Bernasconi and his colleagues published three studies (2010a, 2010b and 2011) in which they focused on the electrophysiological indices of accurate TOJ judgments. They showed that the accuracy of TOJ judgements depends on interhemispheric decoupling at a very early processing stage (40-70 ms post stimulus), probably providing a clear time tag for the first sound, and it is also influenced by the pre-stimulus beta-band activity. However, in these studies, the inter stimulus interval (ISI) varied across participants as it was fitted to the participants' psychophysical threshold. With this procedure, some late components could be significantly attenuated by averaging across participants. In the current study, we tested the formation of the second event by parametrically varying the ISI within the tone pairs. We assumed to find a qualitative change between the responses evoked by grouped sound pairs vs. those processed as two separate events. One possibility that could produce such a qualitative change is based on the notion that two perceptually distinct sounds evoke two separate onset responses. For example, Loveless and his colleagues (1995) recorded magnetoencephalographic responses (MEG) to binaurally delivered tone pairs in a passive paradigm varying the within-pair SOA between 70 and 500 ms. The responses

were dominated by two N100m peaks. The N100m response provides a good estimate of the auditory cortical component of the auditory N1 response recorded with EEG. The auditory N1 is a response to rapid energy change (e.g., an onset response). It peaks ca. 100 ms from the energy change receiving contributions from multiple brain sources (Näätänen, & Picton, 1987; Giard et al., 1994).

Lewandowska and her colleagues (2008) tested ISIs, which are closer to those of the current study. They compared the responses between ISI = 10 ms and ISI = 60 ms to two binaurally presented pure tones of different frequencies. They showed a P2 attenuation in the longer compared to the shorter ISI condition. A possible explanation of this result, is that the P2 overlapped a second distinct N1 response in the ISI = 60 ms condition, which resulted in the observed reduced P2 amplitude. However, because Lewandowska et al's study (2008) used an active paradigm, a possible second N1 effect cannot be disentangled from a smaller P2 response to a perceptually less ambiguous stimulus pair.

In Experiment I, we varied the within-pair ISI between 35 and 75 ms, i.e., around the psychophysical threshold previously measured in our lab by TOJ (Simon and Winkler, 2018). Our test is based on the assumption that when two constituent events are grouped into a single perceptual event, a single N1 response is elicited. In contrast, when the two events are perceived separately, then two successive N1s are elicited. On this basis, we expect a larger ERP difference between the responses to tone pairs of neighboring ISIs at the border between grouping and separation than with the same ISI difference when either both tone pairs are processed as a single event or both are treated as two separate events. We also tested the sensitivity of the threshold to the

statistical context, because an adaptive event formation mechanism has ecological plausibility (Winkler et al., 1990; Hermann et al., 2015). To this end, in Experiment I, we varied the probabilities of the pairs with different within-pair ISIs across different stimulus blocks. If the probabilities of the different ISIs within the sequence influence the electrophysiological boundary in the given condition, then different boundaries will be obtained across blocks. Specifically, adaptation to the most frequently appearing ISI or to the probability-weighted average of the ISIs in the sequence would result in boundaries being close to the frequent ISI within the sequence. Experiment II, provided controls for the effects found in Experiment I. 1) The presence of unequal probabilities of the within-pair ISIs in the sequences could itself have resulted in finding a threshold-like electrophysiological effect. Therefore, Experiment II employed equal ISI probabilities to confirm that the detection of the threshold was not dependent on unequal ISI probabilities. 2) The context effect found in Experiment I was compatible with the notion of adaptation to the most-frequent/weighted-average ISI. By presenting the different ISIs with equal probabilities, we tested whether the context effect reflected the probability-weighted average of the ISIs present in the sequence. Thus, the presence of an electrophysiological boundary in both Experiments would strongly suggest the existence of qualitatively different processing of sound pairs with different short ISIs, and that the processing differences are detectable by the ERP method. Further, finding threshold changes compatible with adaptation to the probability-weighted average would confirm this hypothesis.

2. Experiment I

2.1. Methods

2.1.1. Participants

Twenty-one healthy young adults (age between 20 and 26 years, 17 right-handed, 13 female) were recruited for this experiment. Participants signed an informed agreement prior to the experiment. The study was approved by the United Ethical Review Committee (EPKEB; the institutional review board). Participants received a modest monetary reward. None of them had hearing thresholds higher than 30 dB at any of the 250 Hz, 500 Hz, 2000 Hz or 4000 Hz frequencies, a threshold higher than 15 dB at 1000 Hz, or a between-ear threshold difference higher than 20 dB. The study was performed in the sound-attenuated electrophysiology laboratory of the University of Szeged.

2.1.2. Stimuli and procedure

Sound pairs comprised two 800-Hz pure tones of 10 ms duration, each (1-1 ms linear rise and fall times). The sound pairs were produced by an E-MU 1212M sound card at 78 dB SPL (measured with looped tones) with 850 ms between pair-onsets. They were delivered via Sennheiser HD25-SP II headphones, the first tone to the right and the second to the left ear. There were four stimulus conditions, each delivered in a separate stimulus block of 1500 tone pairs.

Three out of the four conditions delivered a sequence comprising five types of different tone pairs (Table 1). The pairs differed in the within-pair inter-stimulus interval (ISI: 35, 45, 55, 65, and 75 ms). Four pairs were presented with 0.1 probability and one pair with $p = 0.6$ (frequent stimulus). The frequent-stimulus ISI was in separate blocks either 35, 55, or 75 ms (denoted as C35, C55, and C75 from here on). The fourth condition was an oddball sequence with ISI = 55 as the standard ($p = 0.9$) and ISI = 75 as the deviant ($p = 0.1$). The oddball condition was set to test whether the mismatch negativity (MMN; Näätänen et al., 2007) ERP response was elicited by ISI differences in this parameter range. The order of the pairs within the sequences and the condition order were randomized separately for each participant. During the EEG recording, participants watched a self-selected silent movie with subtitles. Each stimulus block (condition) lasted about 22 minutes. Breaks were provided to participants between the stimulus blocks as needed.

Condition \ ISI (ms)	35	45	55	65	75
Condition I (C35)	0.6	0.1	0.1	0.1	0.1
Condition II (C55)	0.1	0.1	0.6	0.1	0.1
Condition III (C75)	0.1	0.1	0.1	0.1	0.6
Condition IV (Oddball)	0	0	0.9	0	0.1

Table 1. Tone-pair (ISI) probabilities in the four experimental conditions.

Spatial Temporal Order Judgment task

The experimental session started with a behavioral threshold measurement. The participant was instructed to judge the order of two short (10 ms, 1-ms rise and 1-ms fall

times) pure tones (800 Hz) by marking whether the pair started at the right or the left ear – trials starting at the left and right ear were presented with equal probability. There was no time pressure on responding. Two practice phases were carried out before the threshold measurement. In the first phase, six ISI = 150 ms pairs were delivered, and after each button press, feedback was provided at the center of a computer screen placed in front of the participant and the next trial started 500 ms after the button press. In the second practice phase, 6 pairs with ISI = 150 ms and another 6 with ISI = 100 ms were delivered in random order, and a summary of the performance was provided after each series. With this procedure, all participants reached a better than 80% performance in three or fewer series (the inclusion criterion was 75%) indicating proper knowledge of the assignment of the response button assignment. The inter-trial interval measured from the response to the previous sound pair was 600-900 ms (uniform probability). The threshold was measured using a three-down-one-up adaptive algorithm that stopped after eight errors (“up” turning points), and the individual’s threshold was calculated as the average of the last six up turning points or three consecutive errors at the same ISI. The initial ISI was 120 ms and the initial step size was 20 ms. The step size was halved after each downward steps (three consecutive correct responses) until the minimum step size of 5 ms was reached. The order of the tones was randomized and the pair was equally likely to start on either side. The threshold measurement lasted for about 2 minutes.

2.1.3. EEG recording

The EEG was recorded with a 32 channel Biosemi DC coupled amplifier (Biosemi, Amsterdam, Netherlands - <https://www.biosemi.com/index.htm>). The placement of the

electrodes followed the International 10/20 system with additional electrodes placed over the two mastoids and the nose. Voltage was recorded with 1024 sampling rate at 24 bits. The built in low pass filter had a 5th order sinc response with the ~3 dB point at 1/5th of the selected sample rate. In the Biosemi EEG system, an amplifier zero reference setup is realized through the CMS (active) and DRL (passive) electrodes positioned on the two sides of the Cz electrode location (<https://www.biosemi.com/faq/cms&drl.htm>). These contacts form a feedback-loop between the electrode montage and the amplifier replacing the traditional ground electrode.

2.1.4. EEG analysis

The MATLAB R2014a software (MathWorks Inc.) was used for analysis with functions from the EEGLAB (Delorme & Makeig, 2004) and ERPLAB toolboxes (Lopez-Calderon & Luck, 2014). The signal was re-referenced to the nose electrode and filtered between 1 and 20 Hz with a finite impulse response filter (Kaiser window = 5.6533). The 1 Hz high pass filter was chosen because we found slow fluctuations in some of the recordings. The 20 Hz low pass value was necessary because of high-frequency noise, which was caused by the headphone. ICA analysis was conducted on the filtered signals and the components related to eye movements were removed based on the recommendation of the ADJUST 1.1.1 software (Mognon et al., 2011). Epochs were generated with 100 pre- and 800 ms post stimulus intervals for each sound pair and baseline corrected to the pre-stimulus interval. Epochs were rejected from further analysis if their amplitude variation exceeded 100 μ V. For each participant, at least 110 trials/condition were retained after artifact rejection. In order to reduce noise, an artificial

fronto-central EEG channel was averaged from the signals measured at Cz, Fz, FC1, and FC2. Statistical analyses were performed on amplitude measures from this channel.

2.1.5. Statistical analysis

For testing the ERP differences between pairs differing by only one ISI step (10 ms), irrespective of the probability of the pair, point-by-point nonparametric *t* tests using 3000 permuted data sets were performed. Only significant effects ($p < .01$) lasting for at least consecutive 20 data points (~20 ms) are reported and interpreted.

The average amplitude measured from the 210-250 ms post-stimulus interval was selected for further statistical analyses, because it was common to all of the intervals in which significant difference between the ERPs elicited by pairs with adjacent ISIs was obtained in the point-by-point *t* tests. The effects of ISI and statistical context were tested by a repeated measures ANOVA (rmANOVA) with the factors CONTEXT ($N = 3$: C35, C55, and C75) and ISI ($N = 5$: 35, 45, 55, 65, and 75 ms). Whenever Mauchly's test of sphericity was significant, the degrees of freedom were adjusted by the Greenhouse-Geisser correction. The alpha level was set to .05. Pairwise differences were post hoc tested using the Bonferroni correction.

In order to follow-up the apparent shift in the effects of ISI on the N1 morphology between contexts C35 and C55 (see Figure 1) an additional rmANOVA was conducted on a subset of the data: CONTEXT ($N = 2$: ISI 35, 55) and ISI ($N = 3$: ISI 45, 55, and 65).

Spearman's rank correlation was employed to test the relationship between the electrophysiological and behavioral data as the latter could only be established on an ordinal scale. The individual electrophysiological boundaries were assessed based on the Chebyshev distance ($d = \max(|p_i - q_i|)$) between ERPs elicited by tone pairs differing in one ISI step; p is the vector of ERP amplitudes measured from the 210-250 ms interval for the shorter-ISI and q for the longer-ISI tone pair of the comparison (difference based estimation).

3. Results

3.1. Electrophysiological results.

The 'deviants minus standard' difference waveforms measured in the oddball condition did not show significant MMN even when the 20-consecutive-point restriction was removed (see Supplementary Figure I). Therefore, the subtractions between tone-pairs with adjacent ISI values are probably not influenced significantly by contributions from MMN.

The differences between the ERPs of stimulus pairs with adjacent ISI values (10 ms step) were significantly different from zero ($p < .01$) for at least 20 consecutive data points in the following comparisons: for 55 vs. 45 ms, between 211 and 247 ms in the C35 condition; for 65 vs. 55 ms, between 202 and 261 ms in the C55 condition; and for 65 vs. 55 ms between 220 and 248 ms in the C75 condition. Figure 1 shows the ERP response to each sound-pair from the artificial fronto-central EEG channel with the scalp topography of the ERP difference between pairs with adjacent ISIs plotted for the 210-250 interval.

The rmANOVA of the 210-250 ms mean ERP amplitudes yielded a significant effect of ISI ($F(4,80) = 27.72$, $MSE = 1.03$, $p < .001$, $\eta^2 = .58$). Post hoc tests revealed that only the difference between 65 and 55 ms was significant (Table 2).

	Mean difference	Standard Error	Adjusted p	Lower Bound of CI95	Upper Bound of CI95
45 – 35	-.133	.133	1	-.551	.285
55 – 45	-.424	.145	.084	-.881	.033
65 – 55	-.717	.169	.004	-1.251	-.184
75 – 65	-.233	.172	1	-.777	.311

Table 2. Mean voltage differences in the 210-250 ms post stimulus interval between the ERPs elicited by pairs with adjacent ISIs (CI95 = 95% confidence interval).

Neither the main effect of CONTEXT, nor the interaction between CONTEXT and ISI reached significance. However, Figure 1 appears to show a shift between the C35 and C55 contexts: whereas in the C35 context, the average ERP elicited by the ISI=55 ms pairs is similar to those elicited by longer-ISI pairs, in the C55 and C75 contexts, the ERP for ISI=55 ms is more similar to those for the shorter-ISI pairs. The rmANOVA conducted on the relevant subset of data yielded a significant $\text{CONTEXT} \times \text{ISI}$ interaction ($F(2,40) = 3.74$, $MSE = .56$, $p = .032$, $\eta^2 = .158$).

3.2. Behavioral results and correlations between the behavioral and ERP threshold measures

The mean TOJ threshold was 36.07 ms ($SD = 23.32$, min. = 5, max. = 71.67). This threshold did not show significant correlation with the electrophysiological threshold: $r(19) = .345$, $p = .12$.

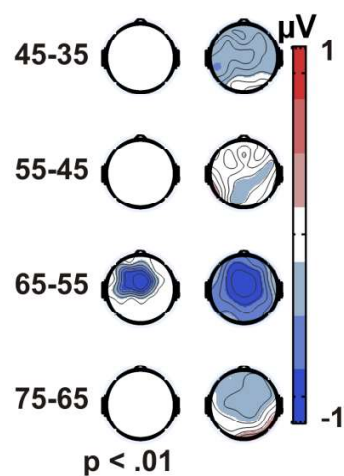
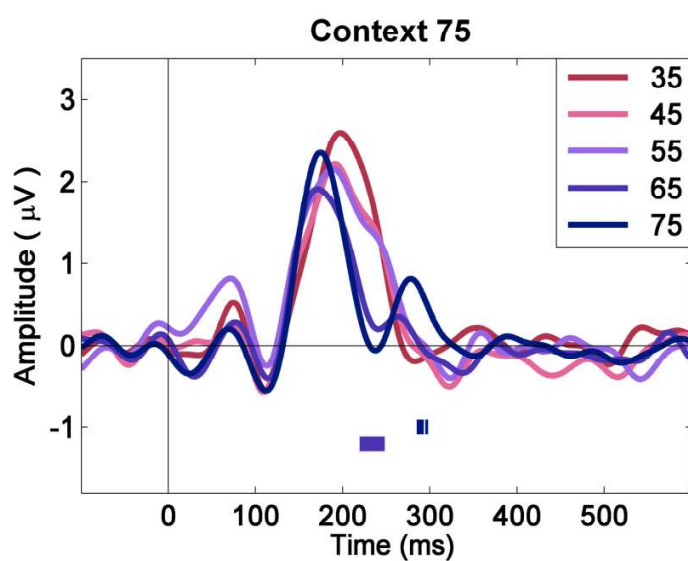
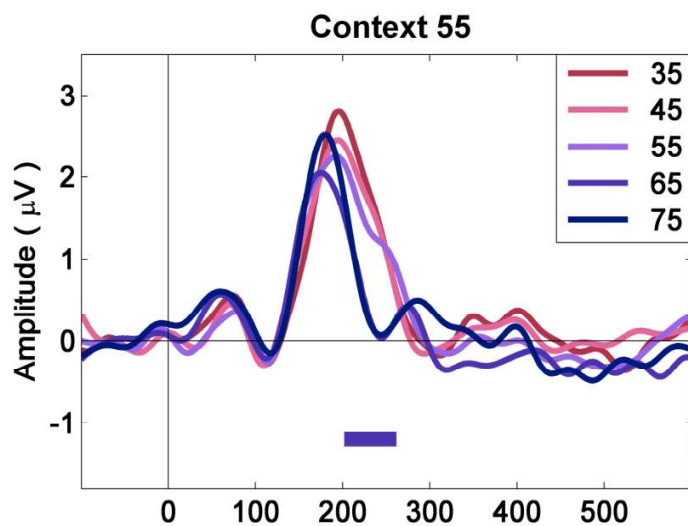
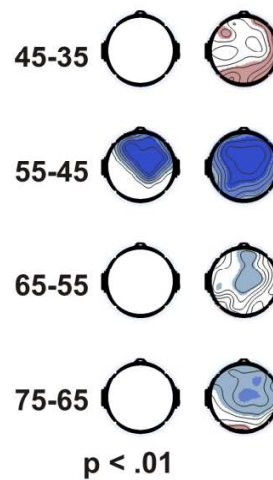
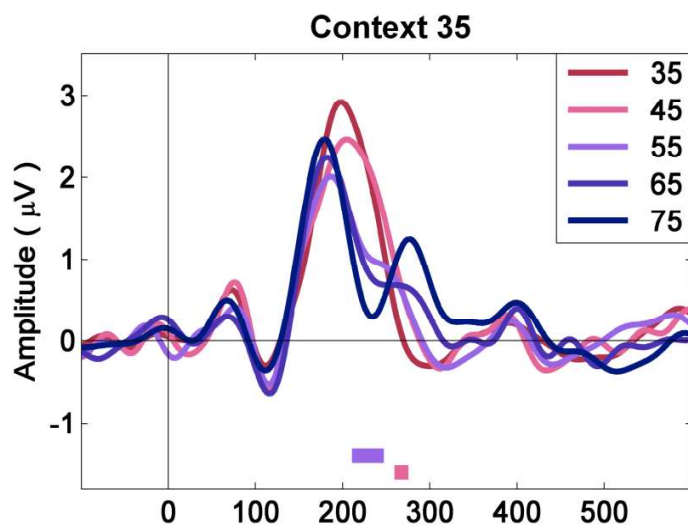


Figure 1. Group-average ($n=21$) ERP responses to individual tone pairs measured from the artificial fronto-central EEG channel. The three panels show the responses for the different frequent-pair conditions: C35, C55 and C75 (top, middle, bottom panel, respectively). In each panel, one response was elicited by the frequent tone-pair of the condition, whereas the other four by infrequent tone-pairs (see Section 3.1) – the lack of MMN in the control condition allows comparing between ERPs elicited by frequent and infrequent pairs similarly as between two infrequent pairs. The horizontal line-segments below the waveforms represent intervals with significant ($p < .01$) point-by-point difference between ERPs elicited by sound pairs with adjacent ISIs (differing in 10 ms); the line color of the horizontal segment corresponds to the color of the ERP waveform representing the pair with the longer of the two ISIs. On the right side of the figure two columns of scalp-distribution displays are shown. The rightmost column depicts the scalp distribution of the ERP amplitude *difference* between pairs with adjacent ISIs measured from the 210-250 ms post-stimulus interval. Cold colors indicate that the pair with the longer of the two ISIs elicited a more negative response than the pair with the shorter ISI. The left column shows the scalp distribution thresholded by the significance of the difference ($p < .01$) between the ERPs elicited by the same two pairs. Note that the thresholding was based on the comparison of average amplitude measurements used in the statistical analysis and thus it does not directly correspond to the point-by-point differences shown below the ERP responses.

4. Experiment II

Although, in the first experiment, we did not find MMN in the control condition, the context manipulation was tied to inequality in the probabilities of the pairs with different

ISIs. This could have affected the finding of an electrophysiological boundary at least in the C35 and C75 conditions in which the frequent ISI was at one or the other end of the range. If the auditory system treated the frequent pair and the nearest one as members of one category (e.g., ISI = 75 ms along with a similar pair with ISI = 65 ms as “long-ISI”) and the rest as the other category (“short-ISI in C75), then the contrast would have resulted in qualitatively differently processing exemplars of the frequent vs. the rare category and thus possibly in the elicitation of ERPs like those obtained in Experiment I. Therefore, in Experiment II, we tested whether the context effect observed in Experiment I was due to unequal stimulus probabilities or to differences in the mean ISI. Two conditions were tested: in the first condition, the ISI varied from 25 ms to 75 ms (six different pairs; Wide ISI range condition), while in the second condition only from 25 ms to 55 ms (four pairs; Narrow ISI range). The order of the first and second condition was not randomized, because the N1 can get attenuated with time (Lange, 2009) decreasing the detectability of a boundary, the primary goal of the second experiment. To assess the effects of time on the position of the boundary, the stimulus blocks (conditions) were a priori split to two halves.

The fixed order of the presentation of the conditions brings up the possible confound that auditory temporal resolution changes due to general arousal/fatigue level between the two conditions as suggested by previous behavioral studies (e.g., Babkoff, Zukerman, Fostick, & Ben-Artzi, 2005). Further, repeated presentations of the stimuli may also influence the ERP response to sound onsets (Ross & Tremblay, 2009). In order to test the possible effects of time on task (time elapsed during the repeated presentation of tone pairs) on the position of the boundary, we separately analyzed the

first and second half of the responses obtained in the two stimulus conditions. If condition half significantly interacted with the ISI variable, it would signal that time on the task affected auditory temporal resolution. Further, we will note the possible effects of the order confound when discussing comparisons between the responses elicited in the two conditions.

4.2. Methods

4.2.1. Participants

Twenty healthy young adults were recruited for this experiment (age between 18-29 years, 17 right-handed, 14 female). They signed an informed consent form and received a modest payment for their participation. The study was approved by the institutional review board (EPKEB). None of the participants had a hearing threshold higher than 35 dB at any of the 250 Hz, 500 Hz, 2000 Hz or 4000 Hz frequencies, or a higher threshold than 20 dB at 1000 Hz or a between ear threshold difference higher than 15 dB at 1000Hz. The study was performed in the sound-attenuated electrophysiology laboratory of the Research Centre for Natural Sciences, Budapest.

4.2.2. Stimuli and procedure

The same stimuli and locations were used as in the first experiment, except that an additional pair (ISI = 25 ms) was introduced and loudness was decreased to 68 dB SPL, because one participant of Experiment I noted that the sounds were quite loud. The sound pairs were produced by a Juli@ MAYA44 sound card (High Quality 24-bit 192 kHz 4x4 reference – ESI Audiotechnik GmbH) and delivered through Sennheiser HD 600 headphones. Each condition delivered 1800 tone pairs, which resulted in 300

pairs/ISI for the wide range condition, and 450 pairs/ISI in the narrow range condition given the equiprobable presentation strategy. Each condition, separately, was delivered as a single continuous stimulus block, which lasted about 25.5 minutes. A break was provided to participants between the two conditions.

Before the EEG recording, there were six threshold measurements: three with a gamified version of the TOJ task and three classic psychophysical measurements, identical to the procedure described for Experiment I. In the gamified version, a red balloon was lifted up or descended corresponding to the performance, thus providing real-time visual feedback. Otherwise the procedure was the same as in the classical version of the task. The gamified measurement always followed the classic measurement after a few minutes break. This way, the classical TOJ measurements are comparable between experiments. Practice was given as described for Experiment I before the first TOJ measurement. One threshold measurement lasted for ca. 2 minutes. The behavioral measurement lasted about 10-12 minutes overall.

Three classical TOJ measurements were used for improving the accuracy of the TOJ measure, whereas the gamified version was introduced as preparation for later on-line studies, to be compared with the classic method.

4.2.3. EEG recording

EEG was recorded with BrainAmp DC 64-channel amplifiers (Brain Products GmbH) and actiCAP active electrodes. A 0.1 Hz high-pass and a 250 Hz low-pass online filters were applied during recording. The placement of the electrodes followed the International 10/20 system. Three additional electrodes were employed: one on the tip

of the nose, one below the left, and the third lateral to the right eye. During recording, the FCz electrode served as the common reference. The EEG was sampled at 1000 Hz with 24 bits resolution.

4.2.4. EEG analysis

The same pipeline of analysis was followed as in Experiment I. After artifact rejection, at least 100 trials were retained for each participant/condition/half/ISI.

4.2.5. Statistical analysis

The difference waveforms were analyzed and electrophysiological boundaries were identified with the same procedure as was described for Experiment I. One exception compared to the procedures in Experiment I is that even significant differences not reaching the 20 consecutive point criterion are shown for the narrow ISI range condition within the latency range corresponding to the significant difference found in the wide ISI range condition.

The interval for amplitude measurements was set to 170-210 ms post stimulus, because the difference effects started earlier in this than in the previous experiment. The effects of the variables were tested by an rmANOVA with the factors of ISI (N = 4: 25, 35, 45, 55), CONDITION (N = 2: ISI: Wide vs. Narrow), and BLOCK HALF (N = 2: first vs. second). ISIs that did not appear in the narrow ISI range condition were not analyzed (but the responses and their differences are shown in Figure 2).

Spearman's rank correlation was conducted between the TOJ and the electrophysiological thresholds.

4.3. Results

4.3.1. Electrophysiological results.

In the Wide ISI range condition, the amplitude difference between the ERPs elicited by stimulus pairs of neighboring ISIs was significant for the '45 – 35 ms' contrast in the 175-218 ms interval. In the Narrow ISI range condition, significant difference was again found for the '45 - 35 ms' contrast. However, this difference appeared only in the 183-192 ms time interval, not meeting the 20 consecutive data point criterion.

The rmANOVA of the mean amplitudes yielded a main effect of BLOCK HALF ($F(1,19) = 10.63$, $MSE = .64$, $p = .004$, $\eta^2 = .35$) and ISI ($F(3,57) = 13.24$, $MSE = .55$, $p < .001$, $\eta^2 = .41$) with no significant interaction between them and very low effect sizes for the interaction effects (effect sizes < 0.06 for the non-significant interactions: $CONDITION \times BLOCK\ HALF$ $\eta^2 = .027$; $CONDITION \times ISI$ $\eta^2 = .028$; $BLOCK\ HALF \times ISI$ $\eta^2 = .057$; $CONDITION \times BLOCK\ HALF \times ISI$ $\eta^2 = .015$) suggesting that time on the task did not influence the amplitude-change as a function of the ISI (although the absolute values of the responses were significantly affected by the time elapsed from the onset of the continuous stimulation). Bonferroni post hoc tests revealed that only the responses to pairs with ISI = 35 and 45 ms differed significantly from each other (Table 3).

	Mean difference	Standard Error	Adjusted p	Lower Bound of CI95	Upper Bound of CI95
35 – 25	.005	.107	1	-.31	.32
45 – 35	-.470	.099	.001	-.760	-.179
55 – 45	-.108	.137	1	-.511	.296

Table 3. Mean voltage differences in the 175-218 ms post stimulus interval between the ERPs elicited by pairs with adjacent ISIs (CI95 = 95% confidence interval).

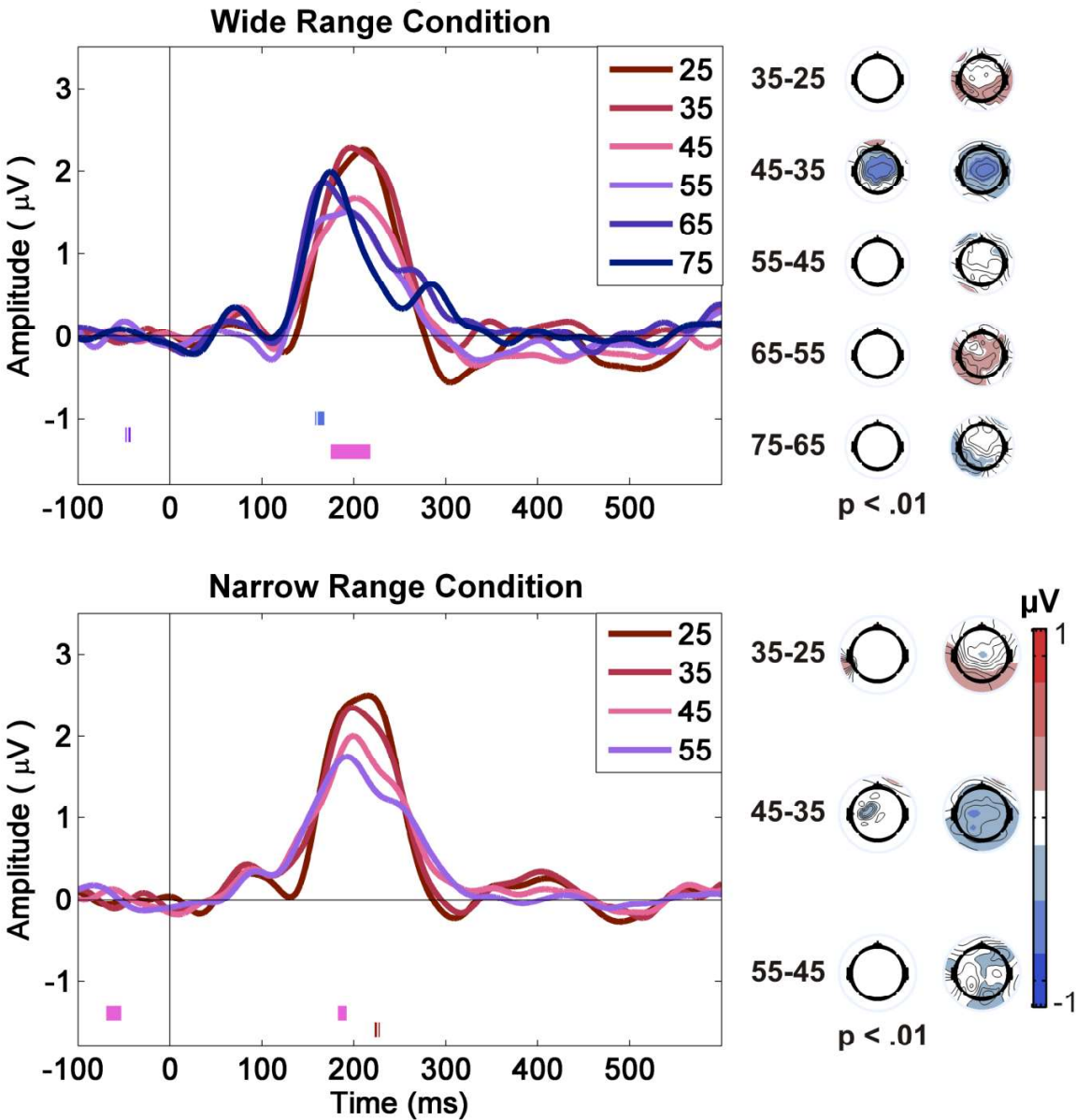


Figure 2. Group-average (n=20) ERP responses to tone pairs measured from the artificial fronto-central EEG channel. The two panels show the responses for the different ISI-range conditions (I: wide, top; II: narrow, bottom). The horizontal line-segments below the waveforms represent intervals with significant ($p < .01$) point-by-

point difference between ERPs elicited by sound pairs with adjacent ISIs (differing in 10 ms); the line color of the horizontal segment corresponds to the color of the ERP waveform representing the pair with the longer of the two ISIs. On the right side of the figure two columns of scalp-distribution displays are shown. The rightmost column depicts the scalp distribution of the ERP amplitude *difference* between pairs with adjacent ISIs measured from the 170-210 ms post-stimulus interval. Cold colors indicate that the pair with the longer of the two ISIs elicited a more negative response than the pair with the shorter ISI. The left column shows the scalp distribution thresholded by the significance of the difference ($p < .01$) between the ERPs elicited by the same two pairs. Note that the thresholding was based on the comparison of average amplitude measurements used in the statistical analysis and thus it does not directly correspond to the point-by-point differences shown below the ERP responses.

4.3.2. Behavioral results

The mean TOJ threshold of the three measurements in the classic psychophysical setup was $M = 37.37$ ms ($SD = 24.96$, min. = 10.83, max. = 87.22). The mean TOJ threshold in the gamified version was $M = 59.64$ ms ($SD = 33.38$, min. = 16.11, max. = 134.44). The results of the two versions of the paradigm significantly differed from each other ($t(19) = -5.877$, $p < .001$, $\eta^2 = .645$), but the correlation between them was strong ($r(18) = .777$, $p < .001$).

No significant correlations were found between the average TOJ and the electrophysiological threshold (Table 4).

	Classic TOJ	Gamified TOJ
--	-------------	--------------

	First block	Whole cond. I	First block	Whole cond. II
rho	.12	.19	.15	.27
p	.601	.40	.507	.24

Table 4. Correlations between electrophysiological indices and behavioral thresholds.

5. General discussion

We found a boundary-like amplitude difference increase in pairwise comparisons between the ERP responses elicited by neighboring sound pairs differing in the time interval between two short dichotic sounds. That is, in each of our stimulus conditions, the amplitude difference was significantly larger between a single pair of neighboring tone-pairs at some non-terminal position of the tested ISI range than that between any other pair of neighboring tone-pairs. This is a sign of a qualitative processing change occurring between these pairs, i.e., a boundary between different processes as opposed to a continuous effect of the within-pair ISI. This pattern of results was expected based on the assumption that with longer ISIs, the second tone would trigger the formation of a second perceptual event. The emergence of a second onset response should cause a larger change between the ERP waveforms characteristic of the fused vs. the split event than that between tone-pairs differing by the same amount of within-pair ISI but not in the number of events.

A similar temporal processing limit was suggested by Steinschneider et al. (1999, 2005) who demonstrated that an additional response component appeared in the Heschl's gyrus at long Voice Onset Times (VOT - the time interval between consonant onset and the onset of the glottal pulsation; VOT determines the difference between voiced and unvoiced consonants). Steinschneider and colleagues obtained similar results also with click trains suggesting the existence of a temporal limit in auditory processing

underlying the categorical perception of voiced and unvoiced consonants. The critical VOT has been found to be a function of sound frequency, showing that this temporal limit is adaptive. Our results are compatible with these findings, and add the observation that this temporal limit is also influenced by the statistical context.

The current results are also compatible with the ‘perceptual moment’ hypothesis (Pöppel, 1971, 1997, 2004; Hirsh & Sherrick, 1961; von Baer, 1864), which defines a temporal limit/frame for *perceptual event* formation. A simple explanation for the non-linear effect of the within-pair ISI is that the larger ERP difference between neighboring tone-pairs is due to the emergence of a second N1-like response above the observed ISI boundary. This explanation does not depend on the existence of a grouping mechanism that would be based on a fixed sampling rate and segment length (such as those proposed to utilize some brain oscillation; see Schwender et al., 1994; Pöppel, 1997)), allowing the temporal limit to change from event to event. The N1 explanation is plausible based on the scalp distribution of the emerging waveform. Studies utilizing sound pairs with short ISIs usually describe a second N1/N100m (Loveless et al., 1995; Fox et al., 2010) but did not test shorter onset differences than 45 ms. Fox et al. (2010) found a second N1 even at SOA = 45 ms (ISI = 25 ms, tone duration = 20 ms) in their adult sample. However, because 45 ms was the shortest onset difference tested in their study, they could not have found the qualitative change observed in the current study. The auditory N1 response has been previously regarded as an exogenous component that is governed by the physical stimulus properties and influenced by the state (e.g., arousal, attention, etc.) of the listener (Näätänen, & Picton, 1987). However, it was later proposed that the N1 is an event-driven rather than stimulus-driven response (Nishihara

et al., 2011). The interpretation that the cause of the boundary-like behavior found in the current study is the emergence of a second N1 at longer within-pair ISIs is compatible with the event-driven conceptualization of the auditory N1: we would then suggest that N1 is elicited when a new sensory event is formed, which leads to the formation of a new perceptual event. The relatively late latency of the significant differences (peaking mostly at >200 ms) and the observed contextual effects are compatible with the notion that the processes underlying the observed threshold may contribute to perception.

Lewandowska et al. (2008) contrasted within-pair SOAs of 20 and 70 ms (ISI = 10/60 ms and duration = 10 ms) and found a large negative difference between the two. These results are also compatible with the current ones. However, Lewandowska and colleagues (2008) interpreted the difference as the suppression of the P2 component elicited by the pair-onset. This is also plausible considering that in their active paradigm, the 70 ms SOA pair was easier to judge than the 20 ms one. The lack of a morphologically separate N1 together with the well discernible P2 component is compatible with this interpretation. The P2 suppression explanation is an example of a more general alternative explanation suggesting a non-linear sensitivity-change with a peak within the tested ISI parameter range. While this may be the case, the elicitation of a second N1 provides a simpler, more plausible explanation.

The lack of reliable correlation between the ERP-based boundary and the one established by behavioral methods prevents us from anchoring the observed electrophysiological boundary to TOJ measurements and through them to the concept of the perceptual moment. One possibility, of course, is that the two measures index different temporal boundaries. Therefore, we cannot confirm that the observed temporal

processing limit matches that suggested by the 'perceptual moment' hypothesis (Pöppel, 1971, 1997, 2004; Hirsh & Sherrick, 1961; von Baer, 1864). However, it is possible that attention (which differed between the EEG and the behavioral measurements) significantly modulates the formation of perceptual events. In support of this hypothesis one should note the well-known attentional sensitivity of the N1 wave (e.g., Hillyard, Hink, Schwent, Picton, 1973). We have putatively linked the N1 with auditory perceptual event formation. Further, the finding of both very short (5 and 10 ms) and long (70 or 80 ms) behavioral thresholds may indicate that participants used different strategies for performing the task. In contrast, the passive ERP measurement is not confounded by listener strategies. It is also possible that the ERP measure benefited from the fixed pair onset-to onset interval, which may have provided a clearer time-tag for the first sound (in an otherwise ambiguous situation) than the random interval used in the TOJ measurement. If this improvement of the TOJ performance reflected reduced threshold for segmenting the two successive sounds, then it could have also affected the electrophysiological threshold. However, if improved tagging improved task performance without changing the threshold, then it does not apply to the current electrophysiological threshold, as it does not rely on order judgement. Finally, the lack of correlation could also have resulted from the noisiness of either or both estimates of the threshold. In addition to the general noise level of the two types of measurements, the ERP based threshold may be sensitive to long-term fluctuations in processing speed as well as to repetition-induced N1 attenuation (Ross & Tremblay, 2009; Schwartze, Farrugia, & Kotz, 2013). The behavioral measure, in turn, could be

sensitive to attentional fluctuations. Better linking the behavioral and electrophysiological findings could be a direction for future studies.

In Experiment I, we found evidence that the probability with which a given ISI appears within the sequence influences the position of the threshold. The effect was not symmetric, as the threshold significantly changed between frequent ISI = 35 and 55 ms (C35 and C55) contexts but not between C55 and C75. Note that the threshold did not follow the distance to the frequent ISI, which would have indicated a mismatch-like response with respect to the expected/prototypical tone-pair. Further, in two of the three contexts (C35 and C75), the threshold-related significant difference was obtained by comparing between responses to pairs with equal (low) probabilities and the scalp distribution of the third significant difference (in the C55 context, where the difference was between the frequent-pair and an infrequent-pair response) was very similar to the scalp distribution obtained for the other two significant differences (Figure 1). Finally, no MMN was obtained in the classic oddball setup (90% standard 10% deviant) or in any other frequent-infrequent comparisons (see Supplementary Figure II). Therefore, it is very unlikely that MMN significantly contributed to the threshold-related differences in Experiment I. A boundary was also observed in Experiment II, in which the probabilities of the different ISIs were equal within the stimulus blocks (conditions). This result further argues against the possibility of an MMN-like source of the ERP effect underlying the observed boundary. One possible explanation of the effect of the statistical context is that the boundary adapts to the mean of the ISI range. However, this assumption is not in line with the lack of change in threshold position between the Narrow and Wide ISI

Range conditions of Experiment II as well as with the position of the threshold in the Wide ISI Range condition (between 35 and 45), which differs from the mean (50 ms).

Admittedly, the fixed condition order of the second experiment allows the possibility that the lack of change in position of the boundary is due to two counteracting factors. For example, the time on task could have increased the boundary while the short ISI dominance could have decreased it. However, this explanation is not supported by the lack of interaction effects (and the very low effects sizes for this non-significant interactions), i.e., that the block halves within conditions did not mediate the ISI effect. Further, because the second of the two conditions did not introduce new stimuli, stimulus novelty should not have affected the responses.

Surprisingly, the mean thresholds differed in the two experiments, being shorter in the second experiment. Experiment I yielded an asymmetrical contextual effect of the ISI: the predominance of pairs with shorter ISIs had a larger effect on the responses to pairs with intermediate ISIs than that of pairs with longer ISIs. Therefore, the shorter threshold in Experiment II can be the result of introducing ISI = 25 ms. Although the adaptation to the mean type of explanation of the context effect has been ruled out, a non-linear version of it suggesting higher weights for the short ISIs would explain the current results. A second explanation for the threshold difference between the two experiments can be based on the notion that the equiprobable presentation in Experiment II created a less predictable context than that in Experiment I (i.e., there was no frequent pair that could have been predicted). The less stable context could have forced the auditory system to work at a higher temporal resolution, which could lead to integrating shorter information chunks than in the more predictable situation.

Both explanations are compatible with the lack of difference in the behavioral threshold measure in the two experiments (36 and 37 ms), which were obtained using the same procedure at the beginning of the experimental sessions.

In conclusion, using an ERP method, we have identified a temporal boundary in processing sounds presented in short succession (in the range of a few tens of milliseconds). This boundary likely corresponds to some strong integration process of the acoustic input, although we found no direct correlation with behavioral measures of the most likely candidate, the perceptual moment. We further found that the boundary is not fixed. Rather, it adapts to the immediate acoustical context, the statistical make-up of the sound sequence. The context effect on the boundary does not follow an adaptation toward the unweighted mean, as short intervals appear to have larger effect on it than long intervals.

Funding: This work was supported by the Hungarian Academy of Sciences' Lendület project (LP2012-36/2012).

References

von Baer, K. E. v. (1864). Welche Auffassung der lebenden Natur ist die richtige? Und wie ist diese Auffassung auf die Entomologie anzuwenden? [Which approach to the living nature is the proper one? And how is this approach applied in

entomology]. Reden, Bd. 1 (pp. 237–284). St. Petersburg, Russia: H. Schmitzdorff.

Babkoff, H., Zukerman, G., Fostick, L., Ben-Artzi, E. (2005). Effect of the diurnal rhythm and 24 h of sleep deprivation on dichotic temporal order judgment. *Journal of Sleep Research*. 14: 7-15. DOI: 10.1111/j.1365-2869.2004.00423.x

Bernasconi, F., Manuel, A., Murray, M., & Spierer, L. (2011). Pre-stimulus beta oscillations within left posterior sylvian regions impact auditory temporal order judgment accuracy. *International Journal of Psychophysiology*, 79(2), 244–248. doi:10.1016/j.ijpsycho.2010.10.017

Bernasconi, F., Grivel, J., Murray, M., & Spierer, L. (2010a). Plastic brain mechanisms for attaining auditory temporal order judgment proficiency. *NeuroImage*, 50(3), 1271–1279. doi:10.1016/j.neuroimage.2010.01.016

Bernasconi, F., Grivel, J., Murray, M., & Spierer, L. (2010b). Interhemispheric coupling between the posterior sylvian regions impacts successful auditory temporal order judgment. *Neuropsychologia*, 48(9), 2579–2585. doi:10.1016/j.neuropsychologia.2010.05.004

Binder, M. (2015). Neural correlates of audiovisual temporal processing – Comparison of temporal order and simultaneity judgments. *Neuroscience*, 300, 432–447. doi:10.1016/j.neuroscience.2015.05.011

Cecere, R., Rees, G., & Romei, V. (2015). Individual Differences in Alpha Frequency Drive Crossmodal Illusory Perception. *Current Biology*, 25(2), 231–235. doi:10.1016/j.cub.2014.11.034

- Cowan, N., (1984). On short and long auditory stores. *Psychological Bulletin*, 96, 341-370.
- Cutting, J. E. (1976) Auditory and linguistic process in speech perception: inferences from six fusions in dichotic listening. *Psychol Rev* 83:114 - 140
- Delorme, A. & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods* 134:9-21
- Dent, M. L., & Dooling, R. J. (2004). The Precedence Effect in Three Species of Birds (*Melospittacus undulatus*, *Serinus canaria*, and *Taeniopygia guttata*). *Journal of Comparative Psychology*, 118(3), 325-331. <http://dx.doi.org/10.1037/0735-7036.118.3.325>
- Fink, M., Churan, J., & Wittmann, M. (2005). Assessment of auditory temporal-order thresholds 20 - a comparison of different measurement procedures and the influences of age and gender. *Restorative Neurology and Neuroscience*.23(5-6) 281-296. PMID: 16477090.
- Florin, E., Vuvan, D., Peretz, I., & Baillet, S. (2017). Pre-target neural oscillations predict variability in the detection of small pitch changes. *PLOS ONE*, 12(5), e0177836. doi:10.1371/journal.pone.0177836
- Fostick, L., Eshcoly, R., Shtibelman, H., Nehemia, R., & Levi, H. (2014). Efficacy of temporal processing training to improve phonological awareness among dyslexic and normal reading students. *Journal of Experimental Psychology: Human Perception and Performance*, 40(5), 1799. doi:10.1037/a0037527

- 1905
1906
1907 Fostick, L., & Babkoff, H. (2013). Different Response Patterns Between Auditory
1908 Spectral and Spatial Temporal Order Judgment (TOJ). *Experimental*
1909 *Psychology*, 60(6), 432–443. doi:10.1027/1618-3169/a000216
1910
1911
1912
1913 Fox, A.M., Anderson, M., Reid, C., Smith, T. Bishop, D.V.M. (2010). Maturation of
1914 auditory temporal integration and inhibition assessed with event-related
1915 potentials (ERPs). *BMC Neuroscience*, 11, p. 49, 10.1186/1471-2202-11-49
1916
1917
1918
1919
1920 Galambos, R., Makeig, S., & Talmachoff, P. J. (1981). A 40 Hz auditory potential
1921 recorded from the human scalp. *Proceedings of the National Academy of*
1922 *Science USA*, 78, 2643-2647
1923
1924
1925
1926
1927
1928 Giard, M. H., Perrin, F., Echallier, J. F., Thévenet, M., Froment, J. C., & Pernier, J.
1929 (1994). Dissociation of temporal and frontal components in the human auditory
1930 N1 wave: a scalp current density and dipole model
1931 analysis. *Electroencephalography and Clinical Neurophysiology/Evoked*
1932 *Potentials Section*, 92(3), 238–252. doi:10.1016/0168-5597(94)90067-1
1933
1934
1935
1936
1937
1938
1939 Herrmann, B., Henry, M., Fromboluti, E., McAuley, J., & Obleser, J. (2015). Statistical
1940 context shapes stimulus-specific adaptation in human auditory cortex. *Journal of*
1941 *Neurophysiology*, 113(7), 2582–2591. doi:10.1152/jn.00634.2014
1942
1943
1944
1945
1946 Hillyard, S. A., Hink, R., Schwent, V. L., & Picton, T. (1973). Electrical signs of selective
1947 attention in the human brain. *Science*, 162, 177–180.
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960

- Humes, L., Busey, T., Craig, J., & Kewley-Port, D. (2013). Are age-related changes in cognitive function driven by age-related changes in sensory processing? *Attention, Perception, & Psychophysics*, 75(3), 508–524. doi:10.3758/s13414-012-0406-9
- Kristofferson, A.B. (1967). Successiveness discrimination as a two-state, quantal process. *Science* 158, 1337–1339
- Lange, K. (2009). Brain correlates of early auditory processing are attenuated by expectations for time and pitch. *Brain and Cognition*, 69(1), 127–137. doi:10.1016/j.bandc.2008.06.004
- Lewandowska, M., Bekisz, M., Szymaszek, A., Wrobel, A., & Szelag, E. (2008). Towards electrophysiological correlates of auditory perception of temporal order. *Neuroscience Letters*, 437(2), 139–143. doi:10.1016/j.neulet.2008.03.085
- Litovsky, R. Y., Colburn, H. S., Yost, W. A., & Guzman, S. J. (1999). The precedence effect. *Journal of the Acoustical Society of America*, 106, 1633–1654.
- Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in human neuroscience*, 8, 213.
- Love, S., Petrini, K., Cheng, A., & Pollick, F. (2013). A Psychophysical Investigation of Differences between Synchrony and Temporal Order Judgments. *PLoS ONE*, 8(1), e54798. doi:10.1371/journal.pone.0054798
- Loveless N, Levänen S, Jousmäki V, Sams M, Hari R (1996). Temporal integration in auditory sensory memory: Neuromagnetic evidence. *Electroencephalography and clinical Neurophysiology*, 100: 220-228.

- Massaro, D.W., (1975). *Experimental Psychology and Information Processing*. Chicago: Rand McNally.
- Mauk, M. D., & Buonomano, D. V. (2004). The neural basis of temporal processing. *Annual Review of Neuroscience*, 27, 307–340.
- Merzenich, M. M., Jenkins, W. M., Johnston, P., Schreiner, C., Miller, S. L., & Tallal, P. (1996). Temporal processing deficits of language-learning impaired children ameliorated by training. *Science*, 271, 77–81.
- Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, 48(2), 229–240. doi:10.1111/j.1469-8986.2010.01061.x
- Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, 13, 201–233. doi:10.1017/S0140525X00078407
- Näätänen R, Paavilainen P, Rinne T, Alho K (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin Neurophysiol* 118:2544–2590. 10.1016/j.clinph.2007.04.026
- Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, 24, 375–425.

- Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, 125, 826–859. doi:10.1037/0033-2909.125.6.826
- Nishihara, M., Inui, K., Motomura, E., Otsuru, N., Ushida, T., & Kakigi, R. (2011). Auditory N1 as a change-related automatic response. *Neuroscience Research*, 71(2), 145–148. doi:10.1016/j.neures.2011.07.004
- Pedersen, C. B., & Salomon, G. (1977). Temporal integration of acoustic energy. *Acta Oto-Laryngologica*, 83, 417–423. doi:10.3109/00016487709128866
- Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as “asymmetric sampling in time.” *Speech Communication*, 41(1), 245–255. doi:10.1016/S0167-6393(02)00107-3
- Pöppel, E. (2004). Lost in time: A historical frame, elementary processing units and the 3-second window. *Acta Neurobiologiae Experimentalis*, 64, 295–302.
- Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends in Cognitive Sciences Vol. 1 (2)*:56-61.
- Pöppel, E. (1971). Oscillations as possible basis for time perception. *Studium Generale*, 24, 85–107.
- Ross, B., & Tremblay, K. (2009). Stimulus experience modifies auditory neuromagnetic responses in young and older listeners. *Hearing Research*, 248(1-2), 48–59. doi:10.1016/j.heares.2008.11.012

- Saija, J., Andringa, T., Başkent, D., & Akyürek, E. (2014). Temporal integration of consecutive tones into synthetic vowels demonstrates perceptual assembly in audition. *Journal of Experimental Psychology: Human Perception and Performance*, 40(2), 857. doi:10.1037/a0035146
- Schwartz, M., Farrugia, N., & Kotz, S. (2013). Dissociation of formal and temporal predictability in early auditory evoked potentials. *Neuropsychologia*, 51(2), 320–325. doi:10.1016/j.neuropsychologia.2012.09.037
- Schwender, D., Madler, C., Klasing, S., Peter, K., & Poppet, E. (1994). Anesthetic Control of 40-Hz Brain Activity and Implicit Memory. *Consciousness and Cognition* 3, 129-147
- Shallice, T. (1964) The detection of change and the perceptual moment hypothesis. *Br. J. Stat. Psychol.* 17, 113–135
- Shinozaki, N., Yabe, H., Sato, Y., Hiruma, T., Sutoh, T., Matsuoka, T., & Kaneko, S. (2003). Spectrotemporal window of integration of auditory information in the human brain. *Cognitive Brain Research*, 17(3), 563–571. doi:10.1016/S0926-6410(03)00170-8
- Simon, J., & Winkler, I. (2018). The role of temporal integration in auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*. doi:10.1037/xhp0000564
- von Steinbüchel, N. & Poppet, E. (1993). Domains of rehabilitation: a theoretical perspective. *Behavioral Brain Research* 56, 1-10

- Steinschneider M, Volkov IO, Noh MD, Garell PC, Howard III MA (1999). Temporal encoding of the voice onset time (VOT) phonetic parameter by field potentials recorded directly from human auditory cortex. *J Neurophysiol* 82:2346--2357.
- Steinschneider, M., Volkov, I., Fishman, Y., Oya, H., Arezzo, J., & Howard, M. (2005). Intracortical Responses in Human and Monkey Primary Auditory Cortex Support a Temporal Processing Mechanism for Encoding of the Voice Onset Time Phonetic Parameter. *Cerebral Cortex*, 15(2), 170–186. doi:10.1093/cercor/bhh120
- Stroud, J.M. (1956) The fine structure of psychological time. *Information Theory in Psychology* (Quastler, H., ed.), pp. 174–205, Free Press
- Sussman, E., Winkler, I., Ritter, W., Alho, K., & Näätänen, R. (1999). Temporal integration of auditory stimulus deviance as reflected by the mismatch negativity. *Neuroscience Letters*, 264(1–3), 161–164. doi: 10.1016/S0304-3940(99)00214-1
- Szymaszek, A., Sereda, M., Pöppel, E., & Szegel, E. (2009). Individual differences in the perception of temporal order: The effect of age and cognition. *Cognitive Neuropsychology*, 26(2), 135–147. doi:10.1080/02643290802504742
- Tallal, P., Miller, S., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., et al. (1996). Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science*, 271, 81–84.
- Yabe, H., Tervaniemi, M., Sinkkonen, J., Huottilainen, M., Ilmoniemi, R. J., & Näätänen, R. (1998). Temporal window of integration of auditory information in the human brain. *Psychophysiology*, 35, 615–619. doi:10.1017/S0048577298000183

Yabe, H., Tervaniemi, M., Reinikainen, K., 1997. Temporal window of integration revealed by MMN to sound omission. *Neuroreport* 8 (8), 1971–1974.

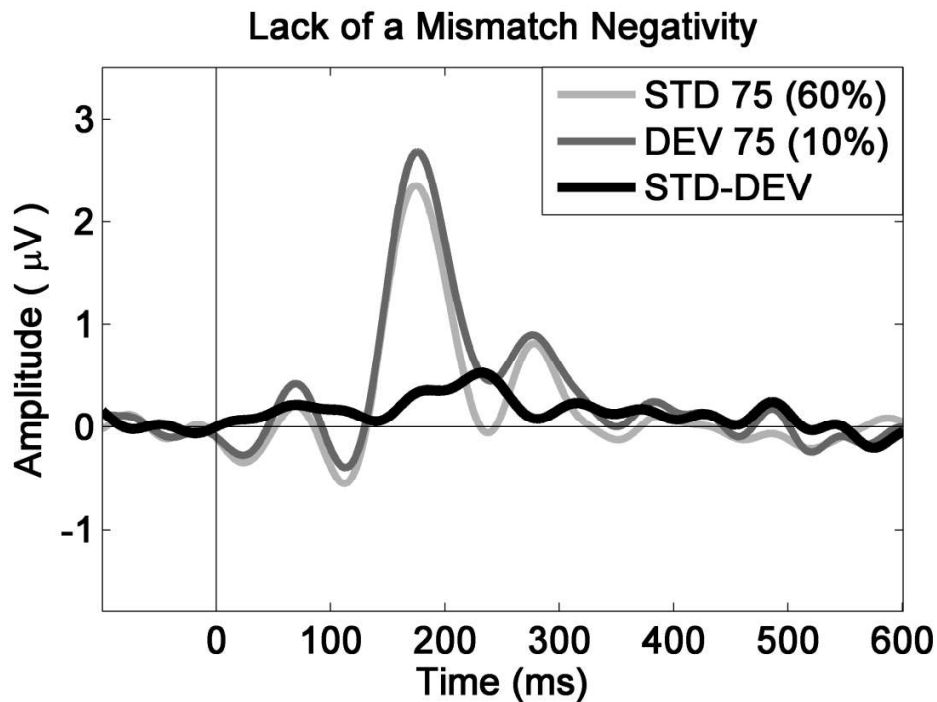
van Eijk, R. L. J., Kohlrausch, A., Juola, J. F., & van de Par, S. (2008). Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type. *Perception & Psychophysics*, 70(6), 955–968.
<https://doi.org/10.3758/PP>

VanRullen, R., & Koch, C. (2003). Is perception discrete or continuous? *Trends in Cognitive Sciences*, 7(5), 207–213. doi:10.1016/S1364-6613(03)00095-0

Winkler, I., Paavilainen, P., Alho, K., Reinikainen, K., Sams, M., & Näätänen, R. (1990). The Effect of Small Variation of the Frequent Auditory Stimulus on the Event-Related Brain Potential to the Infrequent Stimulus. *Psychophysiology*, 27(2), 228–235. doi:10.1111/j.1469-8986.1990.tb00374.x

Zwislocki, J. J. (1960). Theory of temporal auditory summation. *The Journal of the Acoustical Society of America*, 32, 1046–1060. doi: 10.1121/1.1908276

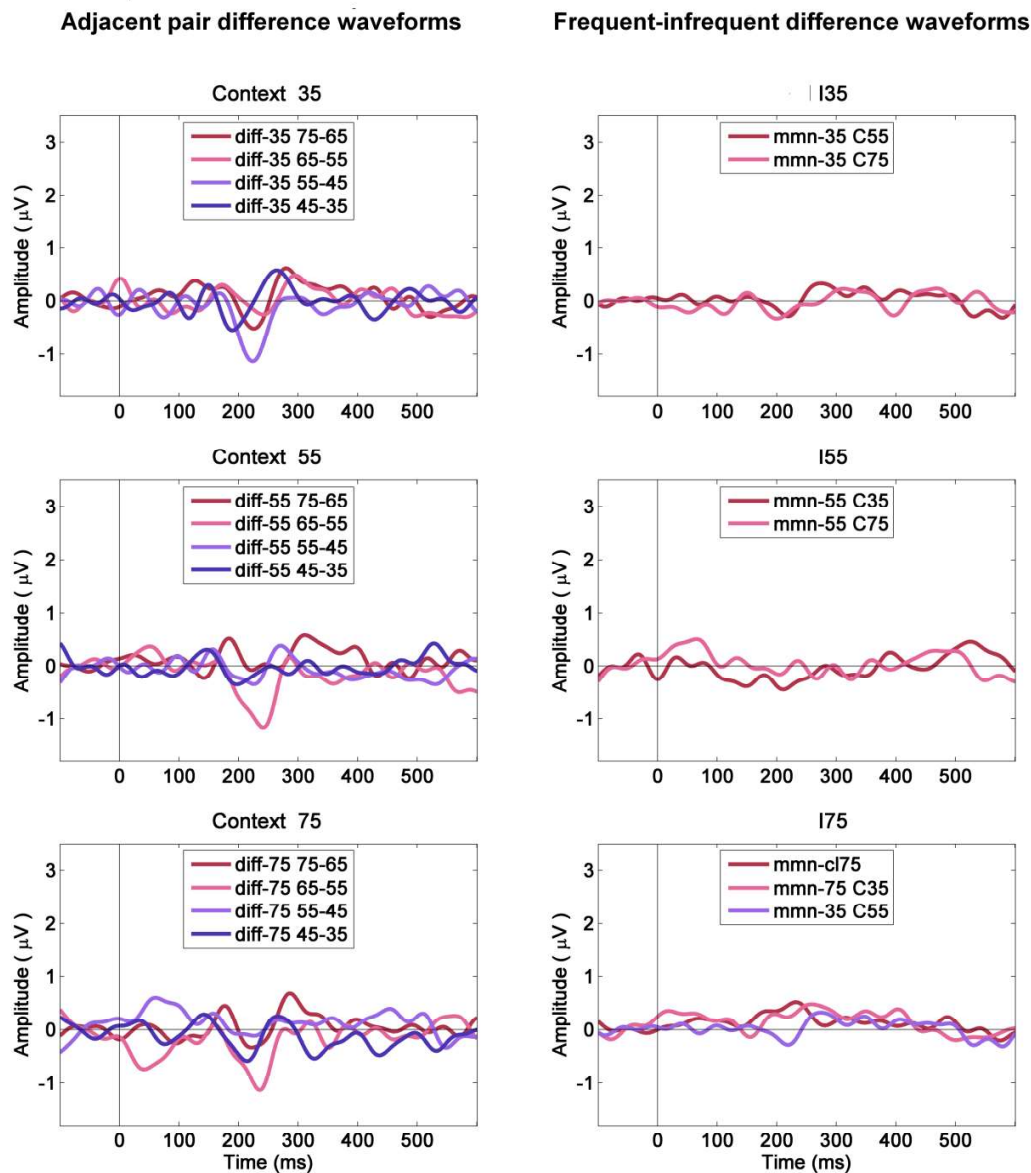
Supplement Figure



Supplement Figure I – The lack of a Mismatch Negativity effect

Group-average ($n=21$) ERP responses to tone pairs measured from the artificial fronto-central EEG channel. The light gray line indicates the ERP elicited by the deviant pair (10% occurrence) with an ISI = 75 ms. The darker gray line represents the ERP of the ISI = 75 ms pair as a standard (60% occurrence). The thick black curve shows the difference waveform. There was no significant difference between the standard and deviant responses (with $p < .01$ criterion). Furthermore, the difference in the region of interest (around 200 ms post stimulus) is positive suggesting that a lack of significant

negativity is not due to a lack of statistical power.



Supplementary Figure II – Difference waves (pairwise and frequent-infrequent comparisons)

The left side of the figure shows the difference waveforms between ERPs elicited by adjacent tone pairs. The right side of the figure shows difference waveforms between ERPs elicited by identical frequent and infrequent tone pairs. Whereas there appears to be an N1-like peak in the difference waveforms between ERPs to adjacent tone pairs, no MMN-like activity appears on the infrequent vs. frequent difference waveforms. The

“mmn-cl75” represents the classical MMN scenario when the tone pair with ISI = 75 ms was presented 10% along with 90% pairs with ISI = 55 ms. The other difference waveforms are taken from the experimental conditions with 60-10-10-10-10% make-up. Panel titles “I35”, “I55” and “I75” refer to the ISI of the tone pairs.